"Toxicogenomics – A Portal for Knowledge Base Development in Toxicology"

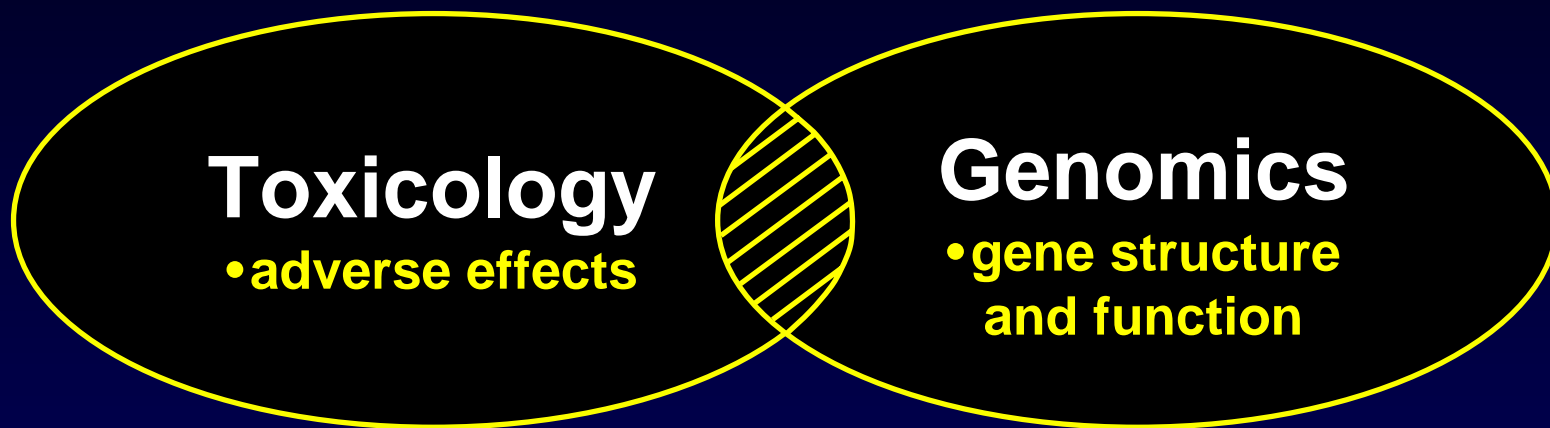## USEPA/ORD Computational Toxicology Workshop
Research Triangle Park, NC
Sept. 29 - 30, 2003

Raymond W. Tennant, Ph.D.
Director, National Center for Toxicogenomics
National Institute of Environmental Health Sciences

NCT
National Center
for Toxicogenomics

NIEHS

NCT

National Center
for Toxicogenomics

Why have an NCT?

The rapid development of genomic technologies provides an unprecedented opportunity to address highly intractable problems of toxicology and environmental health.

➢ **The value of surrogate models for prediction of human health risk**

➢ **Identify biomarkers of incipient adverse effects**

➢ **Harness the results of diverse research efforts for the collective benefit**

➢ **Provide a rational basis for risk assessment**

➢ **Facilitate the identification of specific susceptibility polymorphisms and relate them to environmental diseases**

"…research conducted by the intramural program…has to be second to none and should fulfill a unique mission. It should do those things that are truly inaccessible to extramural institutions, things the nation needs done that neither industry nor academia can do."

Elias Zerhouni

Director, NIH, 2003

Road Map for the Development of Toxicogenomics

Three Objectives

- **Discovery toxicology – identifying and understanding mechanisms of toxicity**

- **Identification of biomarkers of toxicity**

- **Information base development**

NCT
National Center
for Toxicogenomics

- *Discovery toxicology*

  - global gene/protein expression-enhanced identification of mechanisms of adverse effects

  - studies conducted on individual events/processes

Road Map for the Development of Toxicogenomics

Three Objectives

- **Discovery toxicology – identifying and understanding mechanisms of toxicity**

- **Identification of biomarkers of toxicity**
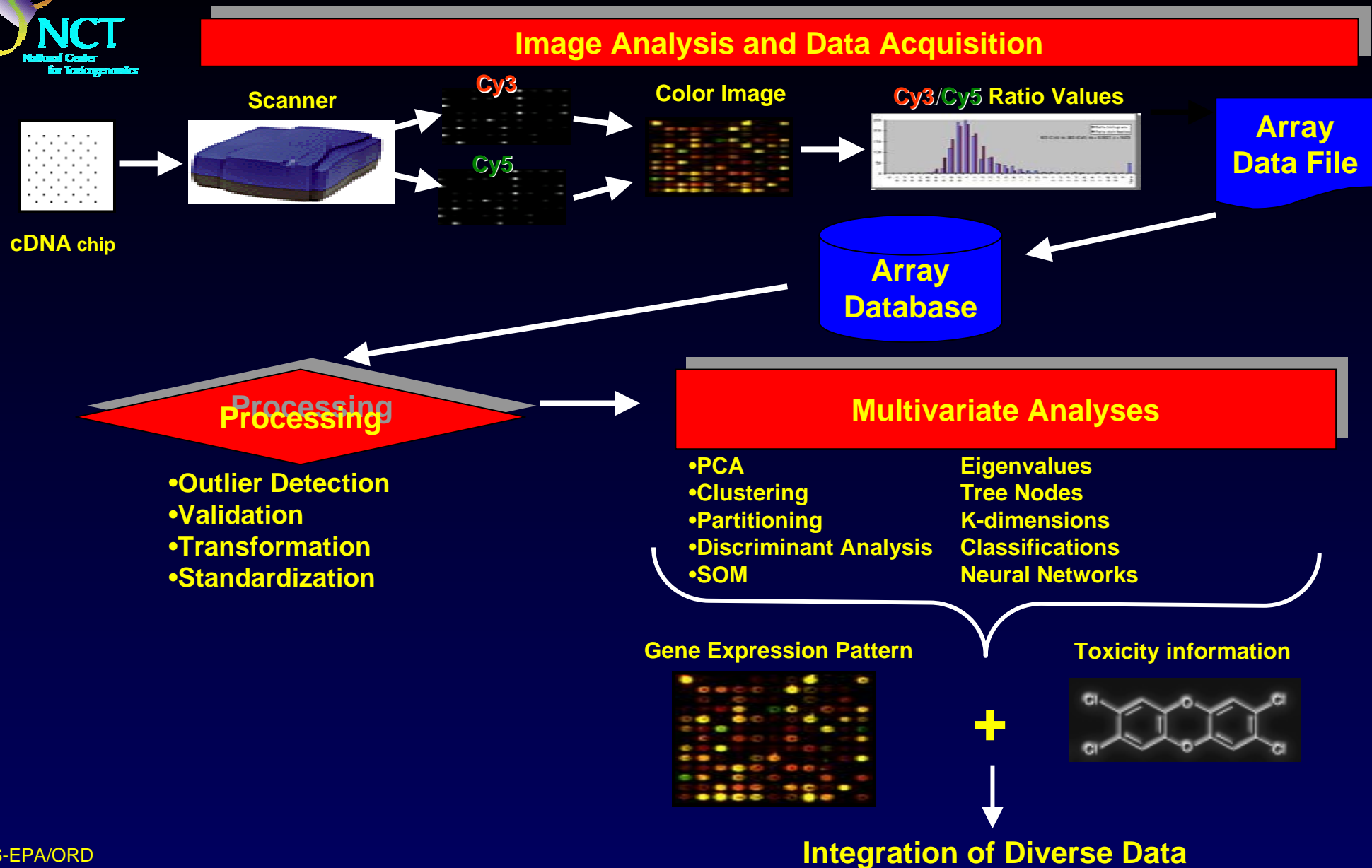
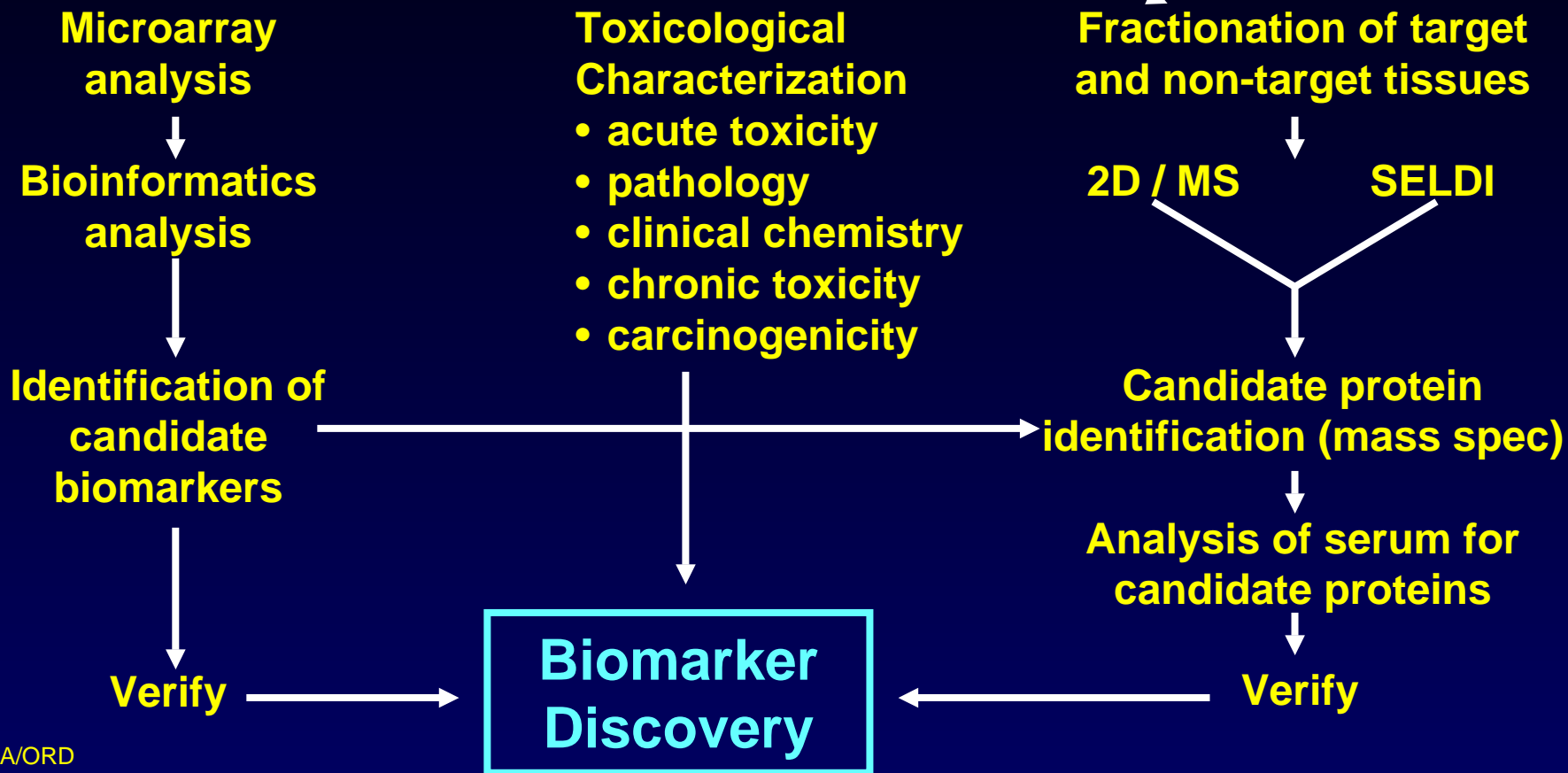- **Information base development**

- *Identification of biomarkers of toxicity*

  - **linkage of altered gene expression patterns to specific adverse effects – "phenotypic anchoring"**

  - **query of gene patterns from multiple agents and effects**

Image Analysis and Data Acquisition

Scanner

Cy3

Cy5

Color Image

Cy3/Cy5 Ratio Values

Array Data File

cDNA chip

Array Database

Processing
- Outlier Detection
- Validation
- Transformation
- Standardization

Multivariate Analyses

- PCA
- Clustering
- Partitioning
- Discriminant Analysis
- SOM

Eigenvalues
Tree Nodes
K-dimensions
Classifications
Neural Networks

Gene Expression Pattern

Toxicity information

+

Integration of Diverse Data

Biomarker Identification Model Based Approach

Chemical exposure of model (dose, time, target site dependent)

Microarray analysis

↓

Bioinformatics analysis

↓

Identification of candidate biomarkers

↓

Verify →

Toxicological Characterization
- acute toxicity
- pathology
- clinical chemistry
- chronic toxicity
- carcinogenicity

Fractionation of target and non-target tissues

↓

2D / MS   **SELDI**

Candidate protein identification (mass spec)

↓

Analysis of serum for candidate proteins

↓

← **Verify**

Biomarker Discovery

Steps to Validate the Toxicogenomic Approach

- **Determine whether gene expression profiles will allow discrimination of compound class**

- **Test the ability to predict unknowns**

NCT
National Center
for Toxicogenomics

Hamadeh, HK, Bushel, PR, Jayadev, S, Martin, K, DiSorbo, O, Sieber, S, Bennett, L, Tennant, RW, Stoll, R, Barrett, JC, Blanchard, K, Paules, RS, and Afshari, CA. (2002) Gene expression analysis reveals chemical-specific profiles. *Toxicological Sciences 67(2): 219-231*.

Hamadeh, HK, Bushel, PR, Jayadev, S, DiSorbo, O, Bennett, L, Li, L, Tennant, R, Stoll, R, Barrett, JC, Paules, RS, Blanchard, K, and Afshari, CA. (2002) Prediction of compound signature using high density gene expression profiling. *Toxicological Sciences 67(2): 232-240*.
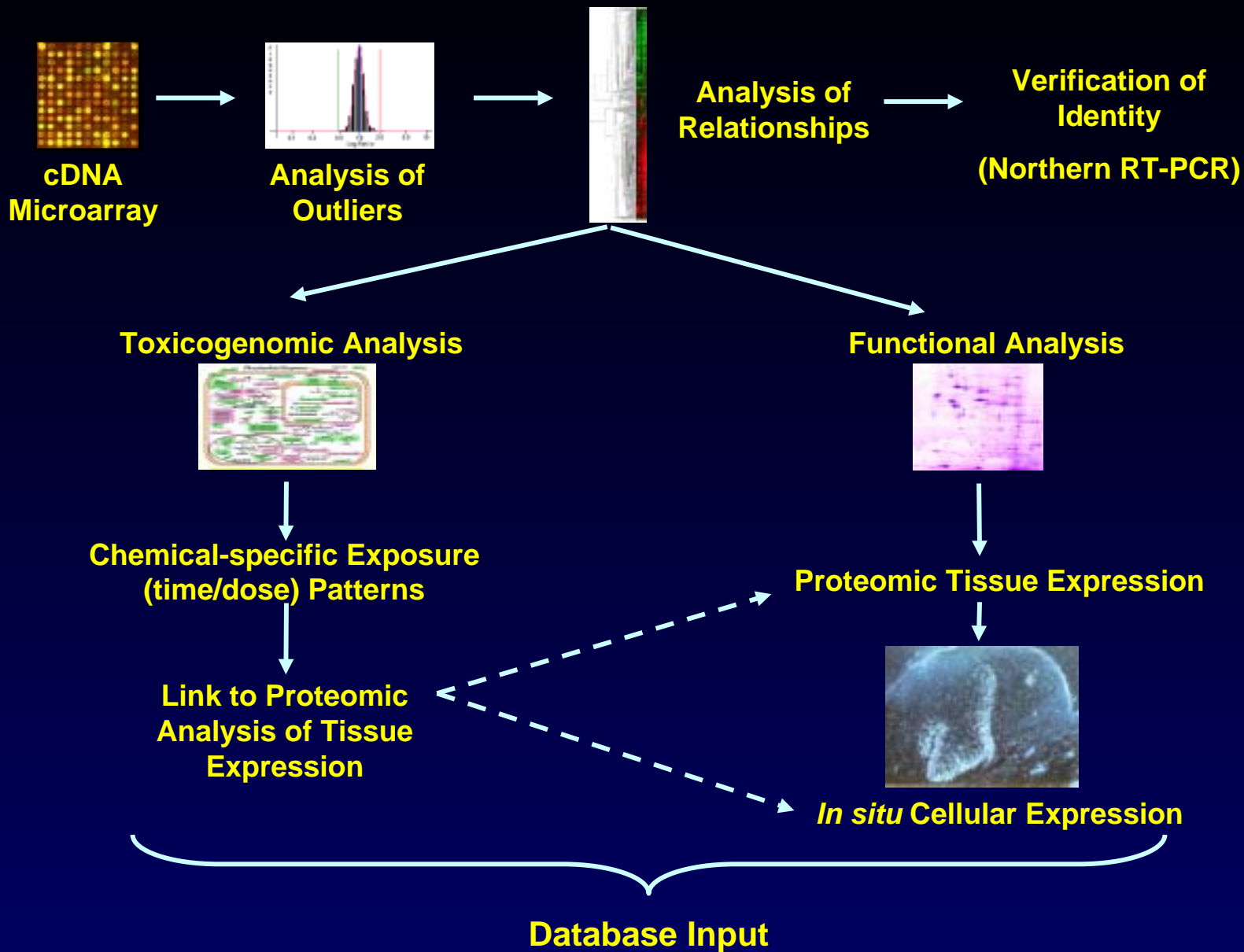
Strategy for Phenotypic Anchoring

- **Profile specific important organ toxicities**
- **Use multiple compounds that elicit that particular toxicity**
- **Use toxic and subtoxic doses and times**
- **Use nontoxic isomers or related compounds if available**
- **Profile early steps in disease processes**
- **Perform analyses at multiple times following exposures**

Hamadeh, H.K., *Knight, B.L., Haugen, A.C., Sieber, S., Amin, R.P., Bushel, P.R., Stoll, R., Blanchard, K., Jayadev, S., Tennant, R.W., Cunningham, M.L., Afshari, C.A. and Paules, R.S. (2002)*
Methapyrilene toxicity: Anchorage of pathological observations to gene expression alterations. *Toxicologic Pathology 30 (4): 470-482.*

Integration of Gene Expression Data

cDNA Microarray → **Analysis of Outliers** → → **Analysis of Relationships** → **Verification of Identity**

(Northern RT-PCR)

Toxicogenomic Analysis

Functional Analysis

Chemical-specific Exposure (time/dose) Patterns

Proteomic Tissue Expression

Link to Proteomic Analysis of Tissue Expression

In situ **Cellular Expression**

Database Input

A Strategy for Toxicogenomics

Short-term Goals

Predictive Assays

- **Signature patterns of exposure**
- **Signature patterns of adverse effects**
- **Proteomic analysis**
- **Biomarker identification**

Long-term Goals

Information Base

- **Gene expression database**
- **Analysis tools (informatics)**
- **Query tools**
- **Relational interfaces and annotation**

Road Map for the Development of Toxicogenomics

Three Objectives

- **Discovery toxicology – identifying and understanding mechanisms of toxicity**

- **Identification of biomarkers of toxicity**

- **Information base development**

What are the Goals of CEBS?

Develop a high quality public knowledge base

- **Create a reference toxicogenomic information system of studies on environmental chemicals/stressors and their effects (a public resource for the scientific community).**

- **Develop relational and descriptive data compendia on toxicologically important genes, groups of genes, SNPs, mutants, and biological phenotypes that are relevant to human health and environmental disease.**

- **Support hypothesis-driven and discovery research in environmental toxicology - and the research needs of risk assessment.**

- *Information base development*

  - **Chemical Effects in Biological Systems (CEBS)**

  Phases

    Data Validity

    Data Aggregation

    Analysis and Query

- *Data Validity*

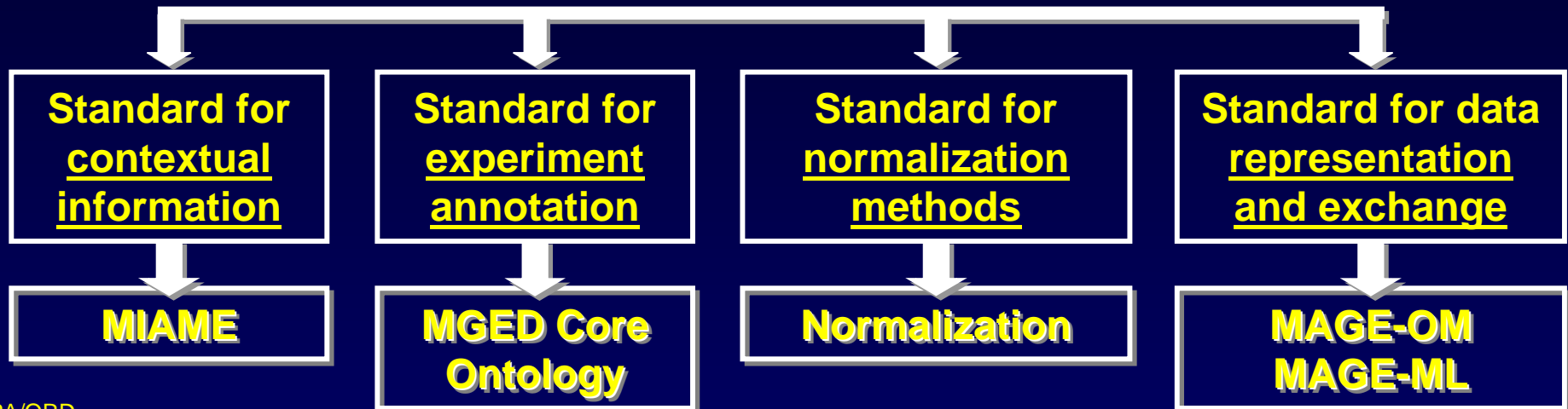**Develop standards for data quality**

MGED Society Standards

- **Microarray Gene Expression Data Society (1999)**

  EBI + major academics + companies in microarray, e.g.

TIGR	Stanford University	Iobion
NCBI	Sanger Centre	Affymetrix
DDBJ	University of California	Agilent / Rosetta

- **MGED Working Groups**

Standard for contextual information	Standard for experiment annotation	Standard for normalization methods	Standard for data representation and exchange
MIAME	MGED Core Ontology	Normalization	MAGE-OM MAGE-ML

MIAME/Tox

- **Define the *core data* common to most experiments**
  - **Minimum/sufficient information**
  - **Structured information**

- **MIAME/Tox is based on MIAME**
  - **<u>Reaffirms</u> the use of MAGE-OM and MAGE-ML standard for data model and exchange format**
  - **<u>Encourages</u> the use of the MGED Ontology for experiment annotation**
  - **<u>Supports</u> MGED standard for recording controls and normalization methods**

- **MIAME/Tox focuses on *tox-specific metadata***
  - **Sample treatments and associated outcomes**

NCT
National Center
for Toxicogenomics

- *Data Aggregation*

  Provide global access via and Oracle database
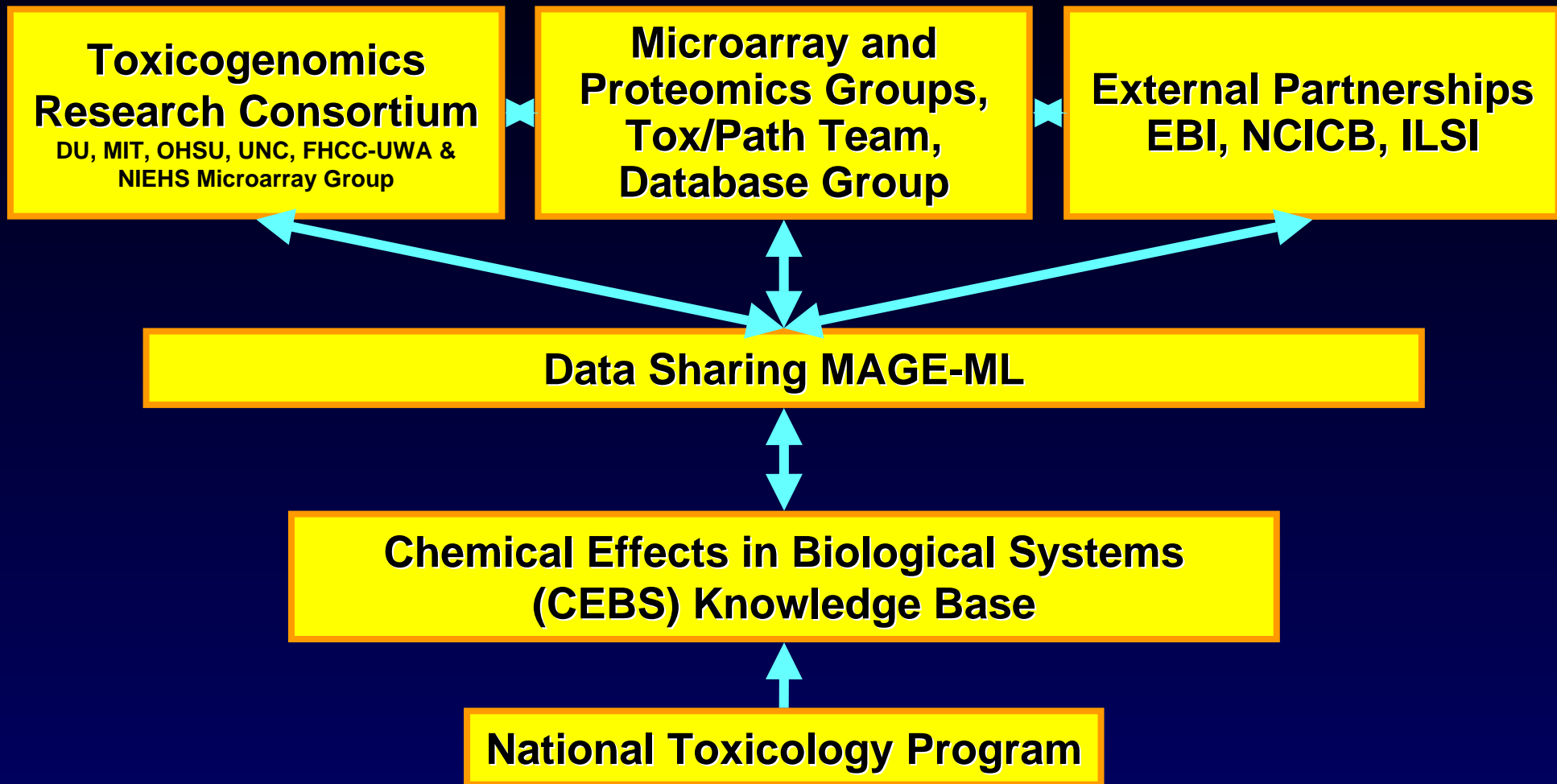
  Seek partners to populate with high quality data

  multi-genome

  multi-agent

  multi- effect

Gaining Content for CEBS

Intramural and Extramural Partnerships

Toxicogenomics Research Consortium
DU, MIT, OHSU, UNC, FHCC-UWA & NIEHS Microarray Group

Microarray and Proteomics Groups, Tox/Path Team, Database Group

External Partnerships EBI, NCICB, ILSI

Data Sharing MAGE-ML

Chemical Effects in Biological Systems (CEBS) Knowledge Base

National Toxicology Program

International Partners

EMBL-EBI

Toxico-genomics

NIEHS-NCT

ILSI-HESI

NIH National Center for Toxicogenomics (NCT)

and

National Toxicology Program

NTP

EMBL-European Bioinformatics Institute (EBI)

and

International Life Sciences Institute (ILSI) Health and Environmental Sciences Institute (HESI)
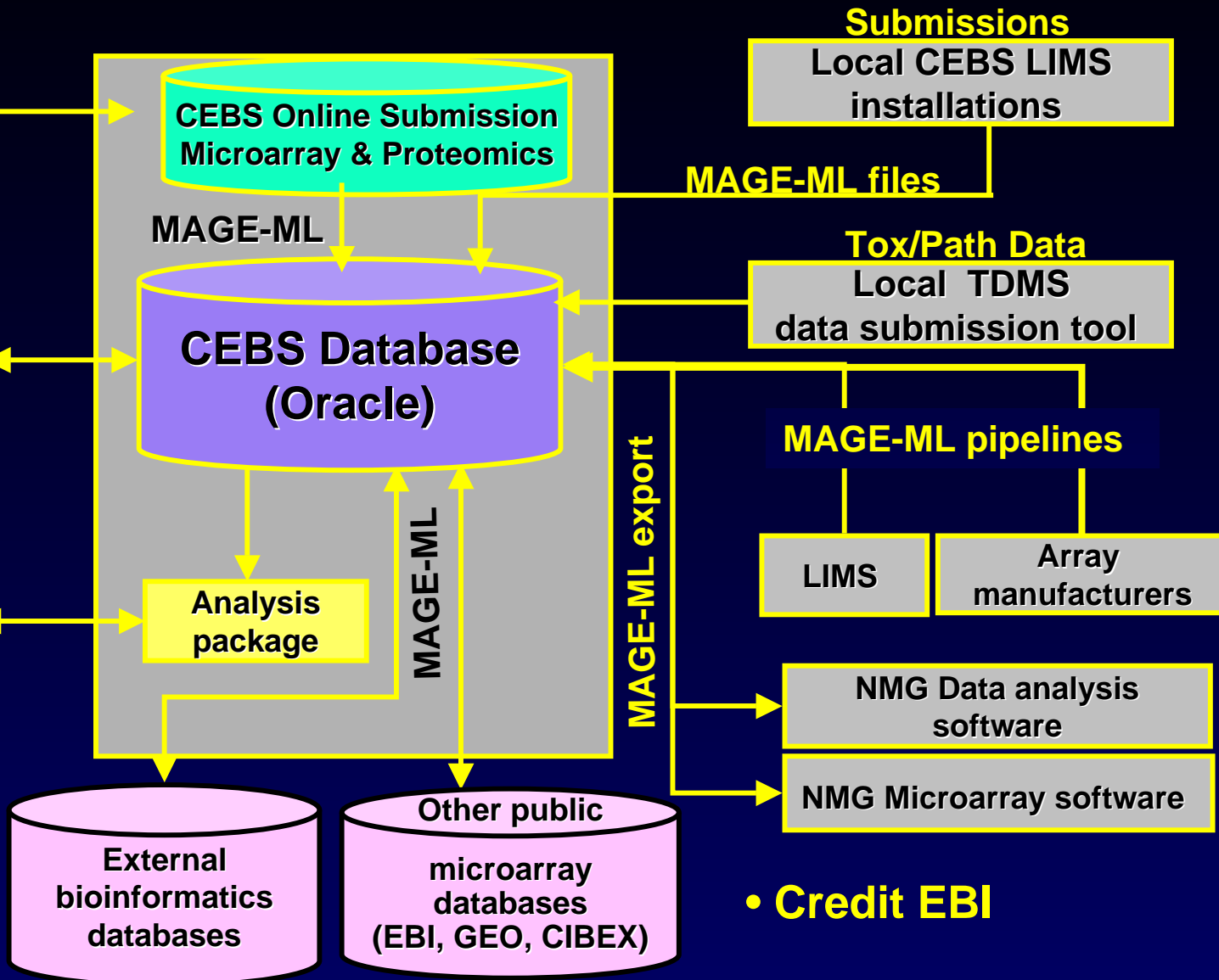
CEBS Infrastructure

C E B S

CHEMICAL

EFFECTS *in*

BIOLOGICAL

SYSTEMS

Home

Microarray

Proteomics

NCT National Center *for* Toxicogenomics

BIOINFORMATICS *to*
KNOWLEDGE

Click on
Microarray to
access the
CEBS
Microarray Site

microarray

MORE INFORMATION

proteomics

MORE INFORMATION

Click on
Proteomics to
access the CEBS
Proteomics Site
(Under
Construction)

Welcome to the National Institute of Environmental Sciences (NIEHS) National Center for Toxicogenomics (NCT) Chemical Effects in Biological Systems (CEBS) knowledge base. The knowledge base is still a work in progress but as we move forward, more and more features will be available to the public and scientific communities.

The NIEHS and NCT are working to help the field of environmental health research evolve into a knowledge-based science in which experimental data are compiled. Computational and informatics tools will play a significant role in improving our understanding of toxicant-related disease. The Chemical Effects in Biological Systems (CEBS) will be a repository for high quality data that is publicly accessible in a relational database that is compatible with standard laboratory output platforms. Database development will be integrated with strategic toxicogenomics experimental design and conduct. Standardized procedures, protocols, data formats, and assessment methods will be used to ensure that data meet a uniform high level of quality. Raw data sets from NCT experiments will be available in their entirety.

Relational and descriptive compendia will be included on toxicologically important genes, groups of genes, SNPs, and mutants and their functional phenotypes. Information about the biological effects of chemicals and other agents and their mechanism of action will be collected from the literature and stored. CEBS will be fully searchable by compound, structure, toxicity, pathology, gene, gene group, SNP, pathway, and network. Dictionaries and explanatory text will guide researchers in understanding toxicogenomics datasets. CEBS will be linked extensively to other databases and to Web genomics and proteomics resources, providing users the suite of information and tools needed to fully interpret toxicogenomics data.

Future promises of CEBS include:

- Developing large context-annotated datasets that allow precise definition of biological/toxicological pathways and lead to the identification of new biomarkers.
- Linking genomic sequence to expression data to determine those genes that may be responsible for the coordinated regulation of sets of genes.
- Increasing the interpretability and dimensionality of expression data by including data from new types of arrays including protein arrays and SNP arrays.
- Aiding in development of new algorithms and computational tools that allow predictive modeling of gene interactions

*Online
Internally at
NIEHS on 18
August 2003*

File   Edit   View   Favorites   Tools   Help

⇦ Back ▾ → ▾ ⊗ ☑ ⌂ | 🔍Search ⭐Favorites 🎦Media ☺ | 🖺▾ 🖨 ✉ | ✂ 🗐 🗐

Address 🔗 http://cebs.niehs.nih.gov/microarray/index.jsp   ▾ ⊘Go

C E B S

CHEMICAL

EFFECTS *in*

BIOLOGICAL

SYSTEMS

NCT National Center *for* Toxicogenomics

BIOINFORMATICS *to*
KNOWLEDGE

LOG IN | REGISTER

SUBMIT A GENE EXPRESSION EXPERIMENT   *go* ▶

Please contribute your tox microarray experiments here and join our community of contributing scientists. Step-by-step instructions for experiment submissions are available online or for download. (Account Required).

Microarray Home

Submit Experiment

Search

Download Center

SEARCH   *go* ▶

Search the CEBS Microarray site for microarray experiments submitted by fellow scientists.
Retrieve detailed experiment information, such as factors studied, protocols, chip design information, etc.
Access raw data files, which are available for download.

TRC Page

ANALYSIS TOOL SUITE   *go* ▶

The CEBS Analysis Tool Suite

First Time Users Please Visit the CEBS Microarray Download Center

ACCESSIBILITY   DISCLAIMER   CREDITS

NIEHS NATIONAL INSTITUTE *of* ENVIRONMENTAL HEALTH SCIENCES

🔵 Internet

- *Information base development*

  - **Chemical Effects in Biological Systems (CEBS)**

  Phases

  Data Validity

  Data Aggregation

  Analysis and Query

Toxicology/Pathology Pathways

- **Capture Domain Knowledge as Toxicology-Pathology-Specific Pathways**

- **Pathway Representation**
  - **BioCarta 320 Pathways currently in CEBS**
  - **KEGG 155 Pathways**
  - **GenMAPP 50 Pathways**
    - **e.g., inflammation**

- **Creation of a Domain Expert Infrastructure**
  - **Web site for Pathway Annotation/Updating**
  - **Committee for Vetting of Annotations/Updates**

Implications of Toxicogenomics for Safety and Risk Assessments – The Why

We rely on simplified assays and models that underestimate the biological complexity underlying toxic effects.

Our current safety and risk assessments are replete with untested, and often unstable, assumptions.

The interfacing of genomics technologies with toxicology provides the most profound way to truly investigate the biological complexity and to create a "Systems Toxicology."

Implications of Toxicogenomics for Safety and Risk Assessments – The What

The capability to observe the dose-rate effects of individual toxicants on the trans-genomic pattern of gene expression provides a capability to understand the full biological complexity underlying adverse effects.

What specific issues in risk assessment can toxicogenomics address?

Specific Issues in Safety and Risk Assessment that can be Addressed by Toxicogenomics

- Provide the capability to identify _new_ mechanisms, pathways, and biomarkers of toxicity.

- Provide the capability to directly assess the relevance of surrogate models for predicting human risk.

- Provide the capability to relate specific susceptibility factors to environmental disease risk.

- Provide the capability to compile data on many potential toxicants, adverse effects, susceptibility factors, etc. in many biological systems and to search for previously unrecognized associations that can lead to new hypotheses of toxicity.

- Provide the capability to understand toxicity and to be able to predict incipient toxicity.

Conclusions

➢ **Genome-based technologies provide an unprecedented opportunity to explore the biological complexity underlying adverse effects.**

➢ **The opportunities are no greater than the challenges of being able to explore high density data.**

➢ **Progress will occur slowly and incrementally; a solid foundation for high quality data assimilation and analysis must be established.**

➢ **The opportunity now exists for creating an information base which compiles the results of diverse individual studies into a compendium of chemical effects in biological systems that can be a resource for the scientific community.**

The Unknown

As we know,

There are known knowns.

There are things we know we know.

We also know

There are known unknowns.

That is to say

We know there are some things

We do not know.

But there are also unknown unknowns,

The ones we don't know we don't know.

Donald H. Rumsfeld

February 12, 2002, Department of Defense news briefing

Acknowledgements

Rick Paules   Microarray Team
Jeff Tucker

Pierre Bushel   Bioinformatics Team

Gary Boorman   ToxPath Team
Rick Irwin
Mike Cunningham

Alex Merrick   Proteomics Team
Ken Tomer

Mike Waters   CEBS Team
Jim Selkirk

Brenda Weis   Toxicogenomics Research
Ben Van Houten   Consortium Team

Stan Stasiewicz   NCT Resource Contract
                                  Management